The tRNA-like domains of *E. coli* and *A. aeolicus* Transfer-messenger RNA: Structural and Functional studies.

Cyril Gaudin[1+], Sylvie Nonin-Lecomte[2+], Carine Tisné[2], Sophie Corvaisier[1], Valérie Bordeau[1], Frédéric Dardel[2*] and Brice Felden[1].

[1]Laboratoire de Biochimie Pharmaceutique UPRES JE2311, Faculté de Pharmacie, Université de Rennes I, 2 avenue du Pr. Léon Bernard, 35043 Rennes, France.

[2]Laboratoire de Cristallographie et RMN Biologiques, UMR8015 CNRS, Faculté de Pharmacie, Université Paris 5, 4, avenue de l'Observatoire, 75006, Paris, France.

[+]The first two authors contributed equally to this work

*Corresponding author : F. Dardel, Cristallographie & RMN Biologiques, Faculté de Pharmacie, 4 avenue de l'Observatoire 75006 Paris, France. Tel: (33) 1 55 73 99 93; Fax: (33) 1 55 73 99 25

e-mail: dardel@pharmacie.univ-paris5.fr

SUMMARY

Transfer-messenger RNA (tmRNA, 10Sa RNA or ssrA) acts to rescue stalled bacterial ribosomes while encoding a peptide tag added trans-translationally to the nascent peptide, targeting it for proteolysis. The understanding at molecular level of this ubiquitous quality control system in eubacteria requires structural information. Here, we describe the purification and structural analysis of a functional fragment of both *A. aeolicus* and *E. coli* tmRNA, recapitulating their tRNA-like domain, which were expressed *in vivo* from synthetic genes. Both recombinant RNA are correctly processed at both 5' and 3' ends and are produced in quantities suitable for structural analysis by NMR and/or X-ray crystallography. The sequence and solution structure of the tRNA-like domains were analysed by various methods including structural mapping with chemical and enzymatic probes and 2D NMR spectroscopy. The minimalist RNAs contain two post-transcriptional base modifications, 5-methyluridine and pseudouridine, as the full-length tmRNA. Both RNAs fold into three stems, a D-analogue, a T-loop and a GAAA tetra-loop. 2D NMR analysis of the imino proton resonances of both RNA allowed the assignment of the three stems and of a number of tertiary interactions. It shows the existence of interactions between the TΨC-loop and the D-analogue, exhibiting a number of similarities and also differences with the canonical tRNA fold, indicating that RNA tertiary interactions can be modulated according to the sequence and secondary structure contexts. Furthermore, the *E. coli* minimalist RNA is aminoacylatable with alanine with a catalytic efficiency an order of magnitude higher than that for full-length tmRNA.

Keywords: tmRNA; structure; RNA probing; NMR; trans-translation

INTRODUCTION

With both tRNA and mRNA properties, transfer-message RNA (tmRNA, SsrA or 10Sa RNA) solves problems that arise when bacterial ribosomes stall during translation [1,2]. It is alanylated at the 3'-CCA tail of its tRNA-like domain prior to entering the stalled ribosome. This alanyl moiety is incorporated into the nascent protein, then translation resumes on a reading frame within tmRNA, and the stalled ribosome is freed at the termination codon. The tmRNA-encoded peptide serves as a signal that triggers degradation of the tagged protein. Recent data suggest that eukaryotes also possess a specific mechanism that eliminates mRNAs lacking termination codons[3].

In all one-piece tmRNA sequences, the 5' end pairs with the 3' end to form an acceptor stem with a 3'-CCA end (Fig. 1) [4]. Native *E. coli* tmRNA contains two modified nucleosides, 5-methyluridine and pseudouridine [5], located within its tRNA-like domain, in a seven-nucleotide loop mimicking the conserved sequence of TΨC loops in canonical tRNAs (T-loop, Fig. 1). The tRNA-like domain of tmRNA also contains an acceptor stem (H1), a D-analogue and a T stem (H6), and is connected by a long disrupted paired region (H5) to a ~ 300 nucleotide-long looped domain (Figure 1) composed of a pseudoknot (PK1), followed by the tag reading frame and a string of three additional pseudoknots (PK2-PK4).

The similarity of the tRNA-like domain of tmRNA to canonical tRNAs raises a number of interesting structural and functional questions : is this domain sufficent for recognition by some of tmRNA protein partners [6-8]: alanyl-tRNA synthetase, elongation factor EF-Tu and SmpB (Small Protein B) ? In the absence of a canonical D-stem and D-loop, what is the tertiary structure organisation of this domain ? In this context, a recent mutagenesis study strongly suggests that an interaction with the TΨC loop and the D-analogue does exist in tmRNA [9]. What are the stabilities of the secondary and the tertiary fold ?

Escherchia coli tmRNA is expected to be shortened without significant loss of function in *trans*-translation, especially within its looped domain, since pseudoknots PK2, PK3 and PK4 are interchangeable and substitutable with single-stranded RNAs [10]. It should just be possible

to construct and analyse the structural and functional properties of shortened versions of tmRNA corresponding to the tRNA-like domain that are still aminoacylatable and able to interact with specific ligands, as SmpB and EF-Tu, but inactive in trans-translation since they are devoid of the tag-encoding sequence.

In the present work, we report the construction of *in vivo* expression system allowing the expression of recombinant RNA fragments recapitulating the tRNA-like domains of *Aquifex aeolicus* and *E. coli* tmRNA. We show that they are processed and modified by *E. coli* hosts cells and that they fold according to the expected secondary structure scheme. NMR solution studies demonstrate the formations of interactions bewteen the TΨC loop and the D-analogue. The tertiary interactions observed share some similarities with canonical tRNA, but also exhibit some interesting differences which could serve as determinants for recognition by specific factors such as SmpB.

RESULTS

Design and construction of tmRNA-TLD

In order to characterise the structural and functional properties of the tRNA like domain of tmRNA, we reasoned that minimal constructs, termed tmRNA-TLD (TLD stands for tRNA-like domain), recapitulating H1, H6, the D-analogue, the T-loop and part of H5, might sustain the tRNA function of tmRNA, *i.e.* its aminoacylation with alanine by alanyl-tRNA synthetase, as well as its ability to interact with small protein B (SmpB) and EF-Tu. Also, the size and overall shape of tmRNA-TLD is reminiscent of that of canonical tRNAs. In this study, two tmRNAs were selected, the *E. coli* one, since it is the molecule for which the largest amount of genetic and biochemical data is available, and its thermophilic *A. aeolicus* counterpart, since the NMR structure of the companion protein SmpB from this organism has recently been solved [11].

An *in vivo* expression system that produces standard tRNAs in quantities suitable for structural and functional studies has been described [12]. This strategy, only applied so far to tRNAs, was extended to tmRNA-TLD. Two constructs were designed, the first one, derived from the *E. coli* tmRNA, contains 61 nucleotides whereas the second one, derived from the *A. aeolicus* sequence, is 60 nucleotide long. For the two constructs, the H5 helix was truncated and capped by a GAAA tetraloop (Figure 1(b-c)). We hypothesised that these engineered, shortened molecules would be recognised and processed by the host enzymes (ribonucleases, methyltransferase and pseudouridine synthase). This seemed a reasonable assumption, given the known data concerning the specificity of these enzymes and our previous experience with several tRNA genes [13,14]. This *in vivo* strategy offers interesting advantages over the *in vitro* transcription methods, especially for structural studies of tmRNA: (i) tmRNA-TLD produced *in vivo* should be processed by the cell enzymes, providing defined functional termini, that are still difficult to obtain by *in vitro* approaches. (ii) The host machinery might also incorporate modified nucleosides post-transcriptionally, which may be required for optimal biological activities and stability of the RNAs.

Therefore, two synthetic genes were constructed with the strategy previously described for tRNA genes, by assembling overlapping oligonucleotides [12,14]. Two recombinant plasmids, pTLD-Ec and pTLD-Ae, were thus obtained as described under Materials and Methods.

Expression and purification of tmRNA-TLD

E. coli strains transformed with the either of the recombinant pTLD plasmids produced a ~60 nucleotide RNA which was clearly visible in an 8M-urea PAGE of total RNA (Figure 1(d)). This band was absent from a total RNA extract from cells carrying the control vector plasmid (not shown). Unambiguous assignment of these RNA species to mature tmRNA-TLD, which were expected to be 61 or 60 nucleotides long, was achieved by quantitative hybridisation to one of the complementary oligonucleotides used for the synthetic gene assembly (not shown). From the relative intensities of the bands in gels revealed by UV shadowing, we estimated that tmRNA-TLD amounts to approximately 5 % of total tRNAs. The purification procedure involved several gel-filtration and ion-exchange chromatography steps, under native conditions, as described under materials and methods. The yields were approximately 5 to 10 mg of purified tmRNA-TLD (85-90% pure) for a 5-litre fermentor culture. An overview of the purification procedure reflecting the various steps is shown in Figure 1(d). Additionally, medium scale preparations of the two ^{15}N labelled TLD were performed as described under material and methods. The final yields after purification were approximately 0.4 mg per litre of labelled medium.

Chemical and enzymatic probing of tmRNA-TLD structure

To demonstrate that the RNA produced and purified *in vivo* were indeed tmRNA-TLD, the RNAs were labelled at their 5' end as described[15], and their adenine and guanine content and location within the RNA sequence was analysed using respectively ribonucleases (RNases) U_2 and T_1 in denaturing conditions. As illustrated in Figure 2, (lanes A_L and G_L), the sequences obtained from the purified RNAs were those expected from the *in vivo* expression of the recombinant tmRNA-TLD gene, thereby confirming the identity of the two purified recombinant RNAs. Next, the prediction of the presence of three stems (H1, H5 and H6) and

three loops (the D-analogue, L5 and L6) was challenged by the use of chemical and enzymatic probes in solution, at 37°C, as described [16]. RNase V_1 cleaves double-stranded (ds) RNA or stacked nucleotides, while nuclease S_1 cleaves single-stranded (ss) RNA. Lead acetate also cleaves ssRNA but with sensitivity to subtle conformational changes of the RNA chain. The reactivity towards these probes was monitored for each nucleotide of both tmRNA-TLDs. Several independent experiments were performed for both purified TLDs (Figure 2a and 2b are representative). The mapping data are summarised in Figure 2 on secondary structure models that they support. For both RNAs, double stranded-specific cuts at nucleotides from H1, H5 and H6 support the existence of these three helices in solution (Figure 2). In *E. coli* TLD, G3 and G4 are both cut by lead, probably because of the G3-U55 wobble pair. For both TLDs, lead and nuclease S_1 cleavages support the presence of loops L5 and L6 capping H5 and H6, respectively. For both RNAs, there are lead cleavages at several nucleotides from the D analogue, consistent with a single stranded RNA portion, but there are exceptions: in *E. coli* TLD, phylogenetically conserved G13 and G14 are cleaved by RNase V_1, suggesting the existence of stacking interactions with other parts of the RNA. At 37°C, however, these interactions might be unstable, since G14 is also cleaved by nuclease S_1 and by lead (Figure 2). In *A. aeolicus* TLD, G12, C17 and A19 are not cleaved by lead. Between H5 and H6, phylogenetically conserved nucleotides G30 and A31 in *A. aeolicus*, (G31 and A32 in *E. coli*) are cut by probes specific of ss RNA, whereas C32 (C33 in *E. coli*) is not.

Solution NMR studies

The solution structures of both *E. coli* and *A. aeolicus* TLDs were investigated by monitoring the imino proton region, which yields information about base-pairing and stacking. The results will be described for *A. aeolicus* TLDs. Differences observed for *E. coli*, if any, will be then pointed out.

2D NMR analyses of the A. aeolicus TLD solution structure

NOESY spectra recorded with 20, 35, 50, 70 and 150 ms mixing times (Fig. 3a and 3b), and $^{1}H[^{15}N]$ SG-TROSY (Figure 3C) were recorded for *A. aeolicus* TLD at 290K, pH 6.5, with no magnesium added.

Identification of two modified nucleotides

The characteristic resonance of a methyl group was unambiguously identified at 1.2 ppm in the 1D and in the 2D NOESY spectra (Fig. 3A), thereby confirming the presence of a modified nucleoside containing a methyl group. Based on earlier work on canonical tRNAs [13,17,18], and full-length tmRNA [5], the presence of a modified nucleoside rT in tmRNA-TLD is quite likely. Here, the 1.2 ppm chemical shift is characteristic of a methyl group attached to a carbon, ruling out N7-methyl guanosine and 2'-OH ribose methylation. From the peak area integration, the level of base modification can be estimated to be close to stoichiometric, within experimental error.

The $^{1}H[^{15}N]$ SG-TROSY shows a characteristic peak resonating at 132 ppm in the nitrogen dimension and 10.9 ppm in proton dimension, which indicates the presence of a pseudouridine (Fig. 3C). This characteristic nitrogen chemical shift is indeed in line with those already reported for the N1 of Ψ on earlier works on several tRNAs [14,17,19-21], and cannot be mistaken for either the N1 of Gs (142-150 ppm) or the N3 of Us, Ts and Ψ□ (around 157 ppm). Therefore, *A. aeolicus* tmRNA-TLD expressed *in vivo* contains both a ribo-thymidine and a pseudouridine.

The T- stem

The NMR identification of the 5-methyl uridine (rT, ribo-thymidine) imino proton mainly relied on the cross-peak with its own methyl (Fig. 3A). The assignment of T38NH was confirmed by a cross-peak pattern similar to those observed in standard tRNAs (Fig. 3A,B and Fig. 4) [17,21,22]. In particular, T38NH and T38CH3 are NOE-connected to one imino proton at 13.02 ppm which belongs to a Watson-Crick G:C base pair, which was assigned to G37.

Imino-imino connections could be followed up to G35 confirming formation of the double-stranded T-stem evidenced by the enzymatic and chemical probing.

The acceptor stem

U54 and G3 form a wobble pair which is the primary determinant for recognition by AlaRS [23,24]. Their imino groups were identified in the HMQC through their particular ^{15}N chemical shifts (Fig. 3C) and also by their strong imino-imino cross-peak which is characteristic of wobble pairing. G3:U54 was used as starting point for the acceptor stem assignment. From NOESY data recorded at different mixing times, connectivities between adjacent Watson-Crick base pairs could be followed from G1:C56 to G5:C52. This confirms the existence of a double-stranded acceptor stem evidenced by the probing experiments.

The anticodon-like stem

The two remaining G resonances at 12.66 and 12.44 ppm are ascribed to two adjacent G:Cs of the truncated anticodon-like arm. The G23NH of the sheared GA pair closing the GAAA loop resonates at 9.32 ppm (not shown).

The TΨC- loop.

In the 2D NOESY experiments, the rT methyl group of tmRNA-TLD gives rise to the classical connectivities observed in canonical TΨC loops of standard tRNAs, ranging from the reverse-Hoogsteen rT54:A58 base pair to pseudouridine 55 and the T–stem on which it is stacked. In particular, in terms of Nuclear Overhauser Effects types and peak intensities, the overall pattern involving T38, G37, Ψ39 and G12 imino protons, and Ψ39 H6 and A48 H8 (Fig. 4), is similar to that of their counterparts T54, G53, Ψ55 and G18 in $tRNA^{Lys}$ [3 21] as explained below:

T38NH chemical shift (14.14 ppm) is consistent with AT base-pair formation. T38 imino does indeed show an NOE to an aromatic proton resonating at 7.95 ppm, which was assigned to A42H8 (Fig. 4). The fact that the tmRNA-TLD is pseudouridylated indeed implies the presence of this A:T reverse-Hoogsteen base pair which is a key recognition element of the *E. coli* TruB pseudouridine synthase, forming several key interactions with a histidine residue of

the enzyme [25]. This T38:A42 base-pair is stacked on G37:C45 as shown by the T38NH-G37NH, T38NH-C45NH_2 and the G37NH-A42H8 cross-peaks (Fig. 3 and 4).

Ψ39N3H shows a strong cross-peak to another imino proton at 11.20 ppm, which is assigned to G12NH on the following grounds : (i) it is a guanosine imino proton, as can be seen on the natural abundance HMQC spectrum. (ii) It pairs with Ψ39, as evidenced by the strong crosspeak between their imino protons (N3H of Ψ55). (iii) The resulting G-Ψ base-pair is stacked onto T38:A42, as evidenced by the numerous crosspeaks (imino-imino, imino-methyl, imino-H8) boxed on figures 3 and 4. This base-pair topology corresponds exactly to what is observed in the tertiary interaction between the TΨC and D loops in canonical tRNAs, where Ψ55 makes a tertiary interaction with G18 in the D-loop. This G18:Ψ55 is stacked on top of T54:A58 which, in turn is stacked onto the T-stem. Thus, most likely, the G paired to Ψ39 in *A. aeolicus* TLD belongs to the D analogue. The two most likely candidates are the evolutionary conserved G12 and G13 [16,26], which mimic the conserved G18 and G19 in canonical tRNAs. By analogy with tRNA structure, we assign the G paired to Ψ39 to G12, as it is the homologue of G18. This is strongly supported by a recent mutational analysis study which suggested a tertiary interaction between the corresponding bases in full-length *E. coli* tmRNA [9].

Interestingly, G12NH and Ψ39N3H yield cross-peaks to a guanosine imino proton at 12.14 ppm (Fig. 3B), this G is engaged into a GC Watson-Crick base-pair with a cytosine, as evidenced by strong GN1H-CN4H_2 crosspeaks (not shown). The most likely candidate for this guanosine is the conserved G13 which would be paired with the conserved C40, mimicking the conserved G19:C56 interaction in standard tRNA. Finally, the second imino proton of Ψ39 (N1H) contacts another GNH at 12.05 ppm (Fig. 3B), which appears not to be engaged in a Watson-Crick interaction. This latter guanosine proton is tentatively assigned to G41, the only remaining guanosine in the TΨC loop. Based on these observations, we thus propose that the stacking within the TΨC loop could be extended by G13:C40, and that G41 could be somewhat looped out of the stack and towards T38.

The conclusion of the NMR studies is that the tmRNA-TLD adopts the cloverleaf like fold with inter-loop interactions and a TΨC loop conformation resembling that of canonical tRNAs, despite the absence of a canonical D stem-loop.

2D NMR analyses of the E. coli TLD solution structure at pH 6.5

Despite some differences, the *E. coli* TLD fold at pH 6.5 resembles that of *A. aeolicus*, and thus the canonical tRNA fold. This study was carried out both without and with $MgCl_2$ (2.5 mM), as it has been shown that magnesium influences the stability and the bending of tmRNA structure [27,28] and thus possibly the inter-loop interaction. Because no major differences were observed between the two conditions apart from small resonance shifts, we will focus on the magnesium-free NOESY spectra.

The T-stem, the acceptor-stem and the anticodon-like stem

The 2D NOESY spectra (Fig. 5A) in the same conditions allowed the identification of the methyl resonance at 1.2 ppm of a rT, which gave crosspeaks to two G imino resonances assigned to G38 and G37 of the T- stem, using the secondary structure model deduced from the probing data. The 1D NOE-difference experiment recorded under the same conditions also disclosed transfers from the methyl resonance to G38 and G37 iminos, and additionally to that of G36 (not shown). Starting from G37 and following imino-imino connections, we were able to assign each of the Watson-Crick G:C pairs of the T-stem ranging from G34:C50 to G38:C46.

All imino protons of the acceptor stem, from G1:C57 to G7:C49, were also assigned through imino-imino cross-peaks, starting from the G3:U55 pair, which was identified as mentioned above for it's *A. aeolicus* counterpart (Fig. 5). The absence of connectivities between the terminal pairs G7:C49 and G34:C50 suggests that little or no stacking occurs between the acceptor and the T stems. As for *A. aeolicus*, the two adjacent G:Cs of the short anticodon-like stem were also identified.

The T- and D- loops

At that stage, the three standard stems evidenced by the probing experiments were clearly assigned (Fig. 2). In the JR-NOESY spectrum recorded at 50ms mixing time (Fig. 5C) two strong imino-imino cross-peaks could be seen, one ascribed above to the G3:U55 pair and the other to a non-canonical base-pair involving one G resonating at 10.2 ppm (as seen on the JR-HMQC, Fig. 5(d)). This base-pair must correspond to a tertiary interaction, since all secondary base-pairs were already assigned. By analogy with both canonical tRNA structures and the *A. aeolicus* TLD (see above), we assigned this base-pair to the G13:(U or Ψ)40, the tertiary interaction demonstrated by Barends and coworkers [9]. This cross-peak is broadened out to baseline at 150ms, indicating fast exchange. This probably explains why we failed to observe the cross-peak corresponding to the N1H of nucleotide 40 in the JR-HMQC experiment. However, as will be shown below, the *E. coli* TLD undergoes a structural change at low pH (≤ 5.0). At pH 4.6, the JR-HMQC clearly shows the Ψ N1 correlation with its characteristic downfield nitrogen shift (not shown). Thus *E. coli* TLD also contains a pseudouridine (Ψ40). The G13:Ψ40 interaction is slightly stabilized by 2.5 mM $MgCl_2$ as shown by the re-appearance of a small cross-peak at 150ms.

Interestingly G13NH and Ψ40NH connect to the same Watson-Crick GNH at 12.5 ppm. As the TΨC loop does not contain any G, and because all Gs of the three stems are already assigned, this last G must belong to the D-loop. According to the 3D models of *E. coli* TLD proposed by the groups of Harvey and Wower [28,29], the only candidate for this interaction is G14 (G19 would be too far away). Moreover, G14 is sequentially in the same position as G13 in *A. aeolicus* TLD, which is engaged in a G13:C40 stacked on the G12:Ψ39 pair. The cross-peak pattern observed here is consistent with a model of TΨC and D-loop interaction in which G13:Ψ40 and G14:C41 are stacked. This is in keeping with the enzymatic probing data (Fig. 2b) which shows that the bond between G13 and G14 is cleaved by ribonuclease V1 which is selective for stacked, base-paired regions.

We were unable to locate T39 imino proton by its NOE cross-peak with the methyl group, even using a 20 ms mixing time, presumably because it exchanges too rapidly with water.

This means that T39 is engaged in a poorly stable base-pair [30]. The imino spectrum showed an unassigned resonance at 13.2 ppm, close to G38NH, yielding a strong cross-peak with water which could correspond to T39NH and suggests a poorly stable A:T39 pair. Moreover, its ^{15}N chemical shift is consistent with either a thymidine or a uridine (Fig. 5(d)). The fact that T39NH possibly resonates very close to G38NH could explain why the sequential cross-peak was not seen.

TLDs solution structure vs. temperature.

Temperature titration does not disclose any fundamental structural transition between 280K and 340K (Fig. 6). The general line-broadening observed at high temperatures is explained by enhanced exchange rates with water. In both TLDs, most stem Watson-Crick G.C pairs are still formed at 340K as well as G12:Ψ39 and A42:T38 in *A. aeolicus*. In *E. coli*, the G13NH resonance, which is involved in the tertiary interaction with Ψ40, is still present at 320K (47°C). The fact that it is not shifted towards 11 ppm as temperature is raised, shows that the broadening to baseline observed at 320K is mainly due to rapid exchange with water and not to a tertiary structure disruption. The temperature shift experiments were repeated for both TLDs in the presence of 2.5 mM $MgCl_2$, and the results were similar, suggesting that the thermal stability of the tRNA-like domain of tmRNAs does not critically depends on magnesium ions.

TLDs solution structure* vs. *pH.

The data obtained with *E. coli* TLD indicate that, at close to neutral pH, its TΨC loop structure is probably slightly more labile than that of its *A. aeolicus* counterpart. The weaker pairings observed for both T39 and Ψ40 precluded observation of some noes from their imino protons. We thus investigated the conformation of both TLDs at lower pH, where the rate of exchange of these resonances with the solvent is decreased.

For *A. aeolicus* TLD, the results were straightforward : the imino proton resonances are largely pH-independent between pH 4.5 and 7 (not shown). A second minor species does start to appear below pH 4.5, showing several downfield-shifted imino protons which could

correspond to N3-protonated cytosines engaged in Hoogsteen C+:G pairs. Nevertheless, all the major cross-peak patterns described above remain unaffected.

For *E. coli* TLD, the experiment was performed in the presence of 2.5 mM $MgCl_2$, in order to improve the stability of the TΨC loop structure. The titration showed that the *E. coli* TLD structure is more pH-sensitive than that of *Aquifex*. (Fig. 7) It clearly shows appearance of a second set of resonances (*i.e.* of an alternate conformation) below pH 6. The relative proportion of the two species, which are in slow exchange at 600 MHz and 293K, is about 1:1 at pH 5.

2D NOESY experiments were recorded on *E. coli* TLD at pH 4.55. Unfortunately, the spectra are complicated by the fact that it is still a mixture of at least two conformations, which prevented obtaining a complete assignment. However, starting from the unambiguous methyl group of T39, a clear pattern of connectivities could be ascribed to the major low-pH form. The T39 imino proton could now be identified at 14.55 ppm (Fig. 7 top). As for Aquifex TLD, the JR-HMQC (not shown) identifies the G-U and G-Ψ pairs, and Ψ40N1H and Ψ40 N3H at respectively 10.9 and 11.5 ppm. The spectrum confirms the presence of T39NH at 14.5 ppm and G13NH at 10.7 ppm. The NOESY spectrum (Fig.7, top) shows that T39 is connected to the G13-Ψ40 pair. It also displays the standard imino-imino crosspeak pattern observed on the Aquifex TLD and in tRNA TΨC -loops linking T39, G13-Ψ40 and G14. An additionnal resonance at 10.4 ppm shows a small crosspeak to T39CH3 (Fig.7, top) and a strong one with G13NH (not shown). Its nitrogen chemical shift probed by the JR-HMQC is that of a uridine. This resonance is tentatively assigned to U12 as suggested by the model of Wower.

Aminoacylation of E. coli *tmRNA-TLD by AlaRS*

Using native gel retardation assays, we collected evidences that labelled tmRNA-TLD interacts with purified alanyl-tRNA synthetase (AlaRS) from *E. coli* (not shown). The enzymatic parameters of *E. coli* tmRNA-TLD alanylation by *E. coli* AlaRS were determined.

Table 1 summarises our comparative kinetic data for aminoacylation of purified *E. coli* tmRNA-TLD and full-length tmRNA with alanine. Interestingly, compared to full-length tmRNA, tmRNA-TLD reaches a two-fold higher plateau, gives a five-fold lower K_M (2 μM) and a two-fold higher k_{cat} (0.5 min^{-1}), suggesting that the domain of tmRNA outside from its tRNA-like portion has negative influence on aminoacylation, regarding both affinity and efficiency.

DISCUSSION

The present work shows that it is possible to express *in vivo* and purify the tRNA-like domain (TLD) of the tmRNA from at least two species, *E. coli* and *A. aeolicus*, in quantities and purity sufficient for structural studies. The yield, approximately 1 mg/litre of culture for both TLD, is however one order of magnitude lower than those previously reported for natural tRNA using the same expression vector. This suggests that processing by *E. coli* enzymes of these non-canonical tRNA-like molecules may be limiting. Both purified RNAs have the expected sequences, as verified by enzymatic sequencing, and are thus correctly processed by *E. coli* ribonucleases on both their 3' and 5' sides. Furthermore, they adopt the predicted secondary structure fold.

Post-transcriptional modifications

The NMR data clearly show the presence of a ribo-thymidine in the TΨC loop of both *E. coli* and *A. aeolicus* TLDs. This methylated base indeed makes numerous NOE contacts with the last base pair in the T-stem, which was unambiguously assigned in both cases. Similarly, it also convincingly proves that U39 is processed into Ψ39 in the *Aquifex* TLD. In the case of the *E. coli* TLD, the more labile TΨC loop structure prevented observation of some of the corresponding imino resonances at neutral pH. However, data collected at low pH show that pseudouridylation does also take place for *E. coli* TLD. The location of the pseudouridine within the TΨC loop of either TLD is again confirmed by numerous NOE contacts with the adjacent pairs, following a standard pattern which is found in canonical tRNAs. These results could be anticipated as both TLDs contain a T-stem and a UUCNNA sequence in the TΨC-loop, the two key recognition elements of *E. coli* pseudouridine synthase [31].

To our knowledge, this is the first time that this recombinant strategy is applied to a non-natural RNA molecule. For NMR studies, incorporation of stable isotopes is a key step for obtaining reliable assignments and structurally relevant information. So far, the only available strategy to produce such labelled RNAs was *in vitro* transcription, using enriched nucleotides triphosphate. However, this technique only allows the synthesis of naked RNA fragments,

devoid of post-trancriptional modifications. Our novel *in vivo* approach allows to circumvent this drawback and thus to investigate the role of modified nucleotides in the structure and the function of such RNAs.

2D and 3D structure characterization

In both TLDs, the prediction of the presence of three stems (H1, H5 and H6) is confirmed by the chemical and enzymatic probing data and by the NMR JR-NOESY spectra. In particular, the two G:U wobble pairs are unambiguously identified. The NMR thermal titrations show that all three stems are very stable, even in the absence of magnesium.

Only the base-paired and/or stack residues could be probed without ambiguity by NMR through their imino proton resonances. The NOESY spectra show that *A. aeolicus* G12 and G13 (G13 and G14 in *E. coli*) are base-paired to Ψ39 and C40 (Ψ 40 and C41 in *E. coli*), which are far from each other in the primary sequence. In the case of *E. coli* TLD, this is in good agreement with our probing data which shows that the G13-G14 phosphodiester bond is sensitive to ribonuclease V1 cleavage. The corresponding bond in *A. aeolicus* TLD (G12-G13) is however not cleaved by V1, although the corresponding G12-Ψ39 and G13-C40 tertiary interactions are supported by the NMR data. The D-analogue of *A. aeolicus* is one base shorter than its *E. coli* counterpart and this could result in a more constrained geometry at the level of the D-TΨC loop interaction. In *E. coli* TLD, there is evidence that the T39:A43 pair contacts an additional unassigned A:U base pair at 13.9 ppm which involves the D-loop and possibly the TΨC loop (Fig. 5). This additional stacking in *E. coli* TLD next to G13 and G14 could provide the structural framework for ribonuclease V1 recognition.

In *E. coli* TLD, the absence of NOE-connectivities between G7:C49, which terminates the acceptor stem, and G34:C50, at the end of the T-stem, suggests that the two helices are not perfectly stacked as they are in canonical tRNAs. This could also result from a more relaxed geometry allowed by the absence of a D-stem and could also partly explain the observations [28] that the inter-stem angle between the acceptor and H5 (110°), is greater than the angles observed between the acceptor branch (acceptor stem and T-stem) and the anticodon stem for any of the known tRNA crystal structures.

Mitochondrial tRNAs often also exhibit large deviations from the canonical cloverleaf geometry, with large variations in the size of the D- and TΨC-loops. Furthermore the conserved nucleotides G18G19 and U54U55C56 within either of these loops are often absent, suggesting that classical tertiary interactions between both domains do not take place [32]. It contrasts with the tRNA portion of tmRNA in which tertiary interactions between the D-analogue and the T-loop mimicking those of canonical tRNAs are present, but in a slightly different way.

Modeling the TΨC-D loop interactions - Comparison with the existing TLD models

Two groups have proposed models of the *E. coli* tmRNA-TLD structure, which were largely based on the homology between tmRNA and canonical tRNAs [28,33]. The present solution studies confirm several of the features of these models for both *E. coli* and *A. aeolicus* TLDs, now providing an accurate picture of the structural mimicry. More specifically, our data convincingly support the existence of the G13:Ψ40 and G14:C41 interactions (*E. coli* TLD numbering) between the D-loop and the TΨC loop. However, in both models, A42 is intercalated in between the G13:Ψ40 and G14:C41, in keeping with the classical tRNA structures (Fig. 8(c)). This is in contradiction with our NMR data which show that there exists two imino-imino contacts between these two base-pairs which should therefore be directly stacked on top of one another, both in the case of the *A. aeolicus* and the *E. coli* TLD. We therefore believe that G41 in *A. aeolicus* TLD is probably at least partially looped out of the structure, possibly closer to the reverse-Hoogsteen A:T base-pair, as suggested by the observation of G41NH_2: Ψ39N1H (not shown) and G41NH:Ψ39N1H (Fig. 3B) cross-peaks. This is further supported by the observation that G41 in *A. aeolicus* TLD is cut by lead and that A42 in *E. coli* TLD is cut by S1.

In the case of *E. coli* TLD, the above observation of an A:U base pair connecting to T39 suggests the possibility of a pairing involving A42, the single remaining adenine in the TΨC loop, since we assumed above that it was not involved in intercalation between G13:Ψ40 and G14:C41. A42 could then pair to an uridine in the D-loop. For instance, either U12 or U17

would be in compatible positions, as suggested by either model. Interestingly, in the *A. aeolicus* sequence, U12 is replaced by an adenine which could account for the difference between the two TLDs. This additional base-pair intercalated between T39 and Ψ40 could be responsible for the partial de-staking of the A43:T39 pair from the T- stem and could also explain the absence of the T39CH3-G38NH cross-peak.

At low pH, however, the *E. coli* cross-peak pattern is consistent with the following stacking: A43:T39/G13:Ψ40/G14:C41 (Fig. 7). The re-appearance of the canonical imino-methyl cross-peaks below pH 5 (Fig. 7) suggests that cytidine protonation (pK 4.01) is responsible for the A43:T39 pair stabilisation. The lowfield part of the spectra also show fast-exchanging protonated-C imino protons consistent with Hoogsteen or reverse-Hoogsteen base-pair formation. Hoogsteen pair formation usually results in backbone stretching (in Z-DNA, adjacent GC+ pairs are 7.7 Å apart from each other). This might lead to disruption of the extra A42:U12/17 pair (the proposed A42:U12/17NH resonance has completely disappeared at pH 4.55), that was assumed to be sandwiched between A43:T39 and G13:Ψ40 at pH 6.5, and consequently to the recovery of a more canonical spatial arrangement, resembling those proposed by Zwieb *et al*. [29] and by Stagg *et al* [28]. The *A. aeolicus* structure would thus be less pH-sensitive because it already has the "good" stack.

Recognition by AlaRS

The major identity determinant for recognition of $tRNA^{Ala}$ by AlaRS is the presence of a G:U wobble pair in the third position starting from the top of the acceptor stem [23,24]. In fact, it has been shown that mini-helix substrates containing this G:U pair can be aminoacylated by the synthetase [34]. Since we showed that the *E. coli* tmRNA-TLD adopted a fold containing both a canonical acceptor and T-stem and contained this major G:U identity element, we expected that it would also be recognised by AlaRS. It was however of interest to compare its aminoacylation parameters to those of the full-length tmRNA in order to dissect the contributions of the various domains of this complex RNA to the recognition by the synthetase.

Overall, compared to full-length tmRNA, tmRNA-TLD has a 9.5-fold gain in catalytic efficiency. Earlier work has reported that full-length synthetic tmRNA binds AlaRS with a 24 μM K_M (to be compared with 9.5 μM using native tmRNA) and a kcat of 2.6 min^{-1} (to be compared with 0.25 min^{-1} using native tmRNA [35]). Between synthetic and native tmRNA, there is a 4-fold difference in the k_{cat}/K_M ratio, in favour of the tmRNA transcript. However, this effect is essentially kinetic, indicating that the activation energy for aminoacylation of the native tmRNA is significantly higher than that of synthetic tmRNA, although the affinity of the latter for the synthetase is lower. Native tmRNA has two post-transcriptional modifications in its TΨC-loop that enhance stability of helices H1 and H6 (B. Felden, unpublished). This could result in increased structural rigidity which could explain the higher activation energy barrier. The gain in catalytic efficiency reported above when tmRNA is reduced to its tRNA-like domain is essentially due to a decrease in K_M, whereas the k_{cat} remains modest compared to that of a naked tmRNA transcript (0.5 min^{-1} *vs*. 2.6 min^{-1}). The higher K_M observed for full length tmRNA compared to tmRNA-TLD can possibly be ascribed to a steric interference by the large pseudoknot/tag domain in the native RNA. However, steric hindrance or the requirement for large conformational changes in the pseudoknot/tag domain cannot explain the poor catalytic efficiency of tmRNA alanylation compared to that of $tRNA^{Ala}$ [35], since alanylation of tmRNA-TLD also proceeds with a similar low k_{cat}. Thus, most likely, it is the specific structure of the D-analogue and its interaction with the TΨC loop which is responsible for the low efficiency of alanylation. This work also suggests that the modified nucleotides present in the TΨC-loop of tmRNA play an important role in the recognition by the synthetase, stabilizing a conformation which is better recognized by the enzyme, as evidenced by the significantly lower K_M of both native tmRNA and tmRNA-TLD, compared with the naked tmRNA transcript. It has indeed been documented that both the rT and the Ψ modifications significantly contribute to the stability of standard tRNA[36]. It is thus reasonable to assume that they play a similar role in the context of the tmRNA.

This "low k_{cat}/normal K_M" profile does make sense in a cellular context, where the recruitment of alanylated-tmRNA by stalled ribosomes is expected to be a much rarer event than the incorporation of alanylated-$tRNA^{Ala}$ in nascent protein chains by active ribosomes. There is thus no need for a high catalytic efficiency of alanylation of tmRNA. However, tmRNA concentration is lower than $tRNA^{Ala}$ (approximately 500 tmRNA/cell compared to 30 000 $tRNA^{Ala}$/cell under exponential growth conditions in *E. coli* [37,38]), there is thus a strong pressure to maintain a high level of affinity, otherwise tmRNA would be out-competed by the "standard substrate" of AlaRS.

The present work shows that our strategy of producing recombinant tmRNA-TLD which are folded and post-transcriptionally modified opens the way to detailed analyses of the function of tmRNA, by allowing the dissection of the respective contributions of the different tmRNA domains. Similar studies could now be undertaken regarding the interaction of tmRNA with either SmpB and/or EF-Tu, looking both at the biochemical parameters and at the possible structural changes induced upon protein binding.

MATERIALS AND METHODS

Construction of the recombinant tmRNA-TLD genes

Construction of the synthetic tmRNA-TLD genes was performed by assembling and ligating synthetic oligonucleotides, as previously described for standard tRNA genes [12]. The resulting double-stranded synthetic DNA fragments were then ligated into the *Eco*RI and *Pst*I sites of the pBSTNAV vector [12]. The DNA sequence of the constructs (plasmids pTLD-Ec and pTLD-Ae) was verified (Cybergene, France).

Expression and purification of tmRNA-TLD

Expression of RNAs from pBSTNAV derived plasmids is constitutive and usually unstable over extended periods of time (F. Dardel, unpublished). For tmRNA-TLD expression, *E. coli* strain JM101Tr was electroporated with purified the purified pTLD plasmids and immediately transferred into flasks of 2xTY medium or Martek-9N (Spetra Stable Isotopes), for ^{15}N labelling supplemented with ampicillin (50 μg/ml).

Large-scale preparation of tmRNA-TLD was similar for both constructs. Briefly, a 5-liter fermentor containing 2xTY medium supplemented with 0.2% glucose (w/v) and 50 μg/ml ampicillin was inoculated with a fresh pre-culture. Cell growth was allowed to proceed under vigorous shaking until it reached the early plateau phase (*i.e.* 3 hours after the end of the log phase), where RNA accumulation is optimal. During growth, the pH was monitored and adjusted to 7.0 by computer-controlled addition of 2N NaOH. For ^{15}N labelling, cells were grown overnight at 37°C in two Fernbach flasks containing each 500 ml Martek-9N (Spectra Stable Isotopes). Cells were harvested by centrifugation and phenol extraction and ethanol precipitation recovered total RNA. Small size RNAs (5S rRNA, tRNAs and tmRNA-TLD) were first separated from bulk nucleic acids by gel filtration on a Superdex-75 prep grade gel filtration column (2.6 x 60 cm; Amersham-Pharmacia Biotech) equilibrated in 50 mM potassium phosphate pH 7.5. Fractions were analysed by 8M urea-PAGE (12% acrylamide)

and checked for the presence of the ~60 nucleotide tmRNA-TLD band (Figure 1(c), lane 2). After this stage, the bulk small size RNAs,were fractionated by ion-exchange chromatography: They were loaded onto a Source-Q column (2.6 x 20 cm; Amersham-Pharmacia Biotech) equilibrated in 50 mM potassium phosphate pH 7.5 and eluted with a 200 to 700 mM NaCl linear gradient. tmRNA-TLD eluted in the tail of the profile, at about 450 mM NaCl (Figure 1(c), lane 3). The corresponding fractions were combined and ethanol-precipitated. After this stage, tmRNA-TLD represented approximately 40 % of the pool. As a final chromatographic step, a second ion-exchange column was performed at a different pH: the RNA pool was re-suspended in 20 mM potassium phosphate, 0.5 mM EDTA, pH 6.5, applied onto the Source-Q column equilibrated in the same buffer and eluted with a 300 to 600 mM NaCl linear gradient. Only the fractions exhibiting the highest purity were pooled. For biochemical studies, tmRNA-TLD were submitted to an additional gel-purification step on a 12% PAGE, followed by electroelution and recovery of purified RNA band by ethanol precipitation.

Enzymatic and chemical RNA probing

Alkaline phosphatase and T4 polynucleotide kinase were from New England Biolabs (Beverly, MA). RNases S1, V1, U2 and T1 were from Amersham-Pharmacia-Biotech (Orsay, France). [γ-^{32}P]ATP (3000 mCi/mmol) was from NEN (Paris, France). Labelling at the 5-' end of the RNA was performed with [γ-^{32}P]ATP and T4 polynucleotide kinase after dephosphorylation by alkaline phosphatase. The RNA was purified on denaturing gel to separate unincorporated [γ-^{32}P]ATP. The appropriate band was eluted passively and the RNA recovered by ethanol precipitation.

Enzymatic digestions and chemical modifications were performed on gel purified 5'- labelled *E.coli* or *A.aeolicus* tmRNA-TLD RNAs (0.5 MBq/reaction), supplemented with 1 μg of total transfer RNA. Labelled RNA was heated to 75°C for 3 min and slowly cooled to room temperature. For both E. coli and *A. aeolicus* TLD, the following amounts of enzymes and chemicals were used : 10 U of nuclease S_1; 10^{-4} U of RNase V_1 and 1,3 mM of lead acetate. For RNA sequencing, 0.2 U of RNase T1 at 0.2 U and U2 at 0.2 U Probing was performed in

a 20-µl-reaction volume at 37°C for 7 minutes [39]. RNAs were ethanol-precipitated, dried and counted and the RNA fragments separated by electrophoresis. Radioactive bands were revealed using a phosphorimager.

NMR spectroscopy

The purified samples were dialyzed several times against 2 mM EDTA water solutions and finally against water, and then freeze-dried. They were resuspended in 95% H_2O-5% D_20 or 99.98 % D_20, 1 mM EDTA solutions. The pH was adjusted at room temperature under mild stirring using 0.1 M HCl and NaOH solutions. No phosphate buffer was added to the samples so as to avoid catalysis of the imino proton exchange that could prevent observation of broad resonances corresponding to weak base-pairs [30]. The *E. coli* tmRNA-TLD sample concentration was 1.5 mM and *A. aeolicus* sample was 1.2 mM. The ^{15}N-labelled samples concentrations were 0.07 mM.

Spectra were recorded at 600 MHz on a Bruker Avance DRX600. The natural abundance ^{15}N-^{1}H GE-JRSE HMQC spectrum [40] of the 61-mer was recorded at 500 Mhz on a Bruker Avance spectrometer equipped with a cryoprobe. In H_2O samples, the water signal was either non-excited using the "Jump and Return (JR)" pulse scheme [41] or suppressed using gradient sculpting [42]. The residual water resonance in D_2O samples was pre-saturated by low-power irradiation during the recycling delay. All data were acquired using oversampling and a 1.3 s repetition delay.

NOESY spectra were recorded in the TPPI mode, and the GE-JRSE HMQC in the States-TPPI or TPPI mode, and the ^{15}N-^{1}H SG-TROSY [43] in the echo-antiecho mode. When necessary, pulse sequences were implemented with carrier frequency jumps so as to acquire spectra on reduced spectral widths. A small 1 ms Z-gradient pulse was applied before the reading JR pulse in JR-NOESY spectra recorded at 110, 130, 200 and 400 ms to cancel out spurious transverse magnetization. For shorter mixing times, (20, 35, 40, 50 and 70 ms), this was achieved using a combination of two small non refocusing Z-gradients. Moreover the first 90° pulse was phase-shifted by 45° so as to avoid the –180 orientation during phase

cycling (Smallcombe, 1993). Mixing times up to 50ms were randomized by 1% to avoid detection of zero-quanta transitions. The natural abundance ^{15}N the GE-JRSE HMQC were recorded overnight with a ^{1}H spectral width of 22 ppm, and a wide ^{15}N spectral width (55 ppm) so as to prevent possible confusion between imino and amino protons. Data were processed either using the Felix 2000 package software (Molecular Simulations, Inc.) or the GIFA program [44].

Aminoacylation by AlaRS

Recombinant alanyl-tRNA synthetase was purified on Ni^{2+}-NTA-agarose (QIAGEN), and purity was confirmed on a 10% Tricine-SDS-polyacrylamide gel. Final protein concentration ranged from 0.5 to 2.5 μM. Assays were performed at 37°C in a medium containing 100 mM Tris-HCl (pH 7.5), 40 mM KCl, 40 mM β-mercapto ethanol, 40 mM $MgCl_2$, 8 mM ATP, 0.2 mg/ml of BSA and 42 μM L-[^{3}H] alanine (152 mCi/mmol). Aliquots were spotted onto 3MM Whatman papers at different times and trichloroacetic acid precipitated. Kinetic parameters (K_M, V_{max} and k_{cat}) were determined from Lineweaver-Burk plots. Aminoacylation plateaus of native full-length tmRNA do not exceed 30%.

ACKNOWLEDGEMENTS

This work was supported by the Human Frontier Science Program Research Grant (RG0291/2000-M 100), a Research Grant entitled “Recherche Fondamentale en Microbiologie et maladies infectieuses” (Institut Pasteur) to B.F, by two “ACI Jeunes Chercheurs” grants from the French Ministry of Research, to both F.D. and B.F and by the “Molécules & Cibles thérapeutiques” program from CNRS and INSERM (to F.D.). C.G. was supported by a Ph.D studentship from the region of Brittany.

The authors gratefully acknowledge Drs. Michel Kochoyan and Christian Roumestand for access to their cryoprobe-equipped spectrometers.

REFERENCES

1. Karzai, A.W.,Roche, E.D. & Sauer, R.T. (2000). The SsrA-SmpB system for protein tagging, directed degradation and ribosome rescue. *Nat. Struct. Biol.* **7**, 449-455.

2. Gillet, R. & Felden, B. (2001). Emerging views on tmRNA-mediated protein tagging and ribosome rescue. *Mol. Microbiol.* **42**, 879-885.

3. Frischmeyer, P.A.,van Hoof, A.,O'Donnell, K.,Guerrerio, A.L.,Parker, R. & Dietz, H.C. (2002). An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. *Science* **295**, 2258-2261.

4. Komine, Y.,Kitabatake, M.,Yokogawa, T.,Nishikawa, K. & Inokuchi, H. (1994). A tRNA-like structure is present in 10Sa RNA, a small stable RNA from Escherichia coli. *Proc. Natl. Acad. Sci. U S A* **91**, 9223-9227.

5. Felden, B.,Hanawa, K.,Atkins, J.F.,Himeno, H.,Muto, A.,Gesteland, R.F.,McCloskey, J.A. & Crain, P.F. (1998). Presence and location of modified nucleotides in Escherichia coli tmRNA: structural mimicry with tRNA acceptor branches. *EMBO J.* **17**, 3188-3196.

6. Karzai, A.W.,Susskind, M.M. & Sauer, R.T. (1999). SmpB, a unique RNA-binding protein essential for the peptide-tagging activity of SsrA (tmRNA). *EMBO J.* **18**, 3793-3799.

7. Rudinger-Thirion, J.,Giege, R. & Felden, B. (1999). Aminoacylated tmRNA from Escherichia coli interacts with prokaryotic elongation factor Tu. *RNA* **5**, 989-992.

8. Barends, S.,Karzai, A.W.,Sauer, R.T.,Wower, J. & Kraal, B. (2001). Simultaneous and functional binding of SmpB and EF-Tu-TP to the alanyl acceptor arm of tmRNA. *J. Mol. Biol.* **314**, 9-21.

9. Barends, S.,Bjork, K.,Gultyaev, A.P.,de Smit, M.H.,Pleij, C.W. & Kraal, B. (2002). Functional evidence for D- and T-loop interactions in tmRNA. *FEBS Lett.* **514**, 78-83.

10. Nameki, N.,Tadaki, T.,Himeno, H. & Muto, A. (2000). Three of four pseudoknots in tmRNA are interchangeable and are substitutable with single-stranded RNAs. *FEBS Lett.* **470**, 345-349.

11. Dong, G., Nowakowski, J. & Hoffman, D.W. (2002). Structure of small protein B: the protein component of the tmRNA-SmpB system for ribosome rescue. *EMBO J.* **21**, 1845-

1854.

12. Meinnel, T.,Mechulam, Y. & Fayat, G. (1988). Fast purification of a functional elongator tRNAmet expressed from a synthetic gene in vivo. *Nucleic Acids Res*. **16**, 8095-8096.
13. Wallis, N.G.,Dardel, F. & Blanquet, S. (1995). Heteronuclear NMR studies of the interactions of 15N-labeled methionine- specific transfer RNAs with methionyl-tRNA transformylase. *Biochemistry* **34**, 7668-7677.
14. Tisne, C.,Rigourd, M.,Marquet, R.,Ehresmann, C. & Dardel, F. (2000). NMR and biochemical characterization of recombinant human tRNA(Lys)3 expressed in Escherichia coli: identification of posttranscriptional nucleotide modifications required for efficient initiation of HIV-1 reverse transcription. *RNA* **6**, 1403-1412.
15. Silberklang, M.,Gillum, A.M. & RajBhandary, U.L. (1977). The use of nuclease P1 in sequence analysis of end group labeled RNA. *Nucleic Acids Res*. **4**, 4091-4108.
16. Hanawa-Suetsugu, K.,Bordeau, V.,Himeno, H.,Muto, A. & Felden, B. (2001). Importance of the conserved nucleotides around the tRNA-like structure of Escherichia coli transfer-messenger RNA for protein tagging. *Nucleic Acids Res*. **29**, 4663-4673.
17. Yan, X.,Xue, H.,Liu, H.,Hang, J.,Wong, J.T. & Zhu, G. (2000). NMR studies of Bacillus subtilis tRNA(Trp) hyperexpressed in Escherichia coli. Assignment of imino proton signals and determination of thermal stability. *J. Biol. Chem*. **275**, 6712-6716.
18. Tisne, C. & Dardel, F. (2002). Optimisation of a Peptide library for screening specific RNA ligands by flow-injection NMR. *Comb. Chem. High Throughput Screen*. **5**, 523-529.
19. Griffey, R.H.,Davis, D.,Yamaizumi, Z.,Nishimura, S.,Bax, A.,Hawkins, B. & Poulter, C.D. (1985). 15N-labeled Escherichia coli tRNAfMet, tRNAGlu, tRNATyr, and tRNAPhe. Double resonance and two-dimensional NMR of N1-labeled pseudouridine. *J. Biol. Chem*. **260**, 9734-9741.
20. Choi, B.S. & Redfield, A.G. (1992). NMR study of nitrogen-15-labeled Escherichia coli valine transfer RNA. *Biochemistry* **31**, 12799-12802.
21. Tisne, C.,Roques, B.P. & Dardel, F. (2001). Heteronuclear NMR studies of the interaction of tRNA(Lys)3 with HIV-1 nucleocapsid protein. *J. Mol. Biol*. **306**, 443-454.
22. Hare, D.R. & Reid, B.R. (1982). Nuclear Overhauser assignment of the imino protons of

the acceptor helix and the ribothymidine helix in the nuclear magnetic resonance spectrum of Escherichia coli isoleucine transfer ribonucleic acid: evidence for costacked helices in solution. *Biochemistry* **21**, 5129-5135.

23. McClain, W.H. & Foss, K. (1988). Changing the identity of a tRNA by introducing a G-U wobble pair near the 3' acceptor end. *Science* **240**, 793-796.
24. Hou, Y.M. & Schimmel, P. (1988). A simple structural feature is a major determinant of the identity of a transfer RNA. *Nature* **333**, 140-145.
25. Hoang, C. & Ferre-D'Amare, A.R. (2001). Cocrystal structure of a tRNA Psi55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme. *Cell* **107**, 929-939.
26. Williams, K.P. (2002). The tmRNA Website: invasion by an intron. *Nucleic Acids Res.* **30**, 179-182.
27. Nameki, N.,Felden, B.,Atkins, J.F.,Gesteland, R.F.,Himeno, H. & Muto, A. (1999). Functional and structural analysis of a pseudoknot upstream of the tag-encoded sequence in E. coli tmRNA. *J. Mol. Biol.* **286**, 733-744.
28. Stagg, S.M.,Frazer-Abel, A.A.,Hagerman, P.J. & Harvey, S.C. (2001). Structural studies of the tRNA domain of tmRNA. *J. Mol. Biol.* **309**, 727-735.
29. Zwieb, C.,Guven, S.A.,Wower, I.K. & Wower, J. (2001). Three-dimensional folding of the tRNA-like domain of Escherichia coli tmRNA. *Biochemistry* **40**, 9587-9595.
30. Nonin, S.,Leroy, J.L. & Guéron, M. (1995). Terminal base pairs of oligonucleotides: imino proton exchange and fraying. *Biochemistry* **34**, 10652-10659.
31. Gu, X.,Yu, M.,Ivanetich, K.M. & Santi, D.V. (1998). Molecular recognition of tRNA by tRNA pseudouridine 55 synthase. *Biochemistry* **37**, 339-343.
32. Helm, M.,Brule, H.,Friede, D.,Giege, R.,Putz, D. & Florentz, C. (2000). Search for characteristic structural features of mammalian mitochondrial tRNAs. *RNA* **6**, 1356-1379.
33. Zwieb, C.,Wower, I. & Wower, J. (1999). Comparative sequence analysis of tmRNA. *Nucleic Acids Res.* **27**, 2063-2071.
34. Francklyn, C. & Schimmel, P. (1989). Aminoacylation of RNA minihelices with alanine. *Nature* **337**, 478-481.

35. Barends, S.,Wower, J. & Kraal, B. (2000). Kinetic parameters for tmRNA binding to alanyl-tRNA synthetase and elongation factor Tu from Escherichia coli. *Biochemistry* **39**, 2652-2658.

36. Derrick, W.B. & Horowitz, J. (1993). Probing structural differences between native and in vitro transcribed Escherichia coli valine transfer RNA: evidence for stable base modification-dependent conformers. *Nucleic Acids Res*. **21**, 4948-4953.

37. Lee, S.Y.,Bailey, S.C. & Apirion, D. (1978). Small stable RNAs from *Escherichia coli*: evidence for the existence of new molecules and for a new ribonucleoprotein particle containing 6S RNA. *J. Bacteriol*. **133**, 1015-1023.

38. Bremer, H. & Dennis, P.P. (1987). Modulation of chemical composition and other parameters of the cell by growth rate. In *Escherichia coli and Salmonella typhimurium: cellular and molecular biology* (Neidhardt, F. C., Ingraham, J. L., Low, K. B., Magasanik, B., Schaechter, M. & Umbarger, H. E., eds.), pp. 1527-1542. American Society for Microbiology, Washington, D.C.

39. Felden, B.,Himeno, H.,Muto, A.,McCutcheon, J.P.,Atkins, J.F. & Gesteland, R.F. (1997). Probing the structure of the Escherichia coli 10Sa RNA (tmRNA). *RNA* **3**, 89-103.

40. Szewczak, A.A.,Kellogg, G.W. & Moore, P.B. (1993). Assignment of NH resonances in nucleic acids using natural abundance 15N-1H correlation spectroscopy with spin-echo and gradient pulses. *FEBS Lett*. **327**, 261-264.

41. Plateau, P. & Guéron, M. (1982). Exchangeable proton NMR without base-line distortion, using new strong-pulse sequences. *J. Am. Chem. Soc*. **104**, 7310-7311.

42. Hwang, T.S. & Shaka, A.J. (1995). Water suppresssion using excitation sculpting with gradients. *J. Magn. Reson. A* **112**, 275-279.

43. Weigelt, J. (1998). Single scan, sensitivity- and gradient-enhanced TROSY for multidimensional NMR experiments. *J. Am. Chem. Soc*. **120**, 10778-10779.

44. Delsuc, M.A. (1989). A new maximum entropy processing algorithm with applications to nuclear magnetic resonance experiments. In *Maximum Entropy and Bayesian Methods* (Skilling, J., ed.), pp. 285-290. Kluwer Academic Publishers, Dordrecht, The Netherlands.

FIGURE LEGENDS

Figure 1.

(a). Secondary structure of tmRNA from *E.coli* with emphasis on its tRNA-like domain (TLD). The looped mRNA domain (PK1-PK4 plus the tag reading frame) is shown schematically. (b) and (c). *E. coli* and *A. aeolicus* tmRNA-TLD sequences, respectively. (d). Purification steps of E. coli-TLD expressed *in vivo*. 12% polyacrylamide/ 8M urea gel. Visualisation by UV-shadowing. Lane 1 : bulk RNA after cell lysis and phenol extraction ; lane 2 : « small size » RNAs after gel-filtration ; lane 3 : pH 7,5 ion exchange pool ; lane 4 : pH 6.5 ion exchange pool.

Figure 2.

(a). Partial RNA sequencing and structural probing of *A. aeolicus* TLD (left panel) and E.coli TLD (right panel). Autoradiograms of cleavage products of 5'-labelled RNAs by lead, and nucleases S_1 and V_1. Lane C, incubation controls ; Lane AH, sodium bicarbonate hydrolysis ; Lane G_L, RNase T_1 hydrolysis ladder ; Lane A_L, RNase U_2 hydrolysis ladder. Sequences of both RNAs are indicated on the side. Digestions with various ribonucleases and lead (II) acetate were performed on 5'-labelled tmRNA-TLD supplemented with 1 μg of total transfer RNA. (b). Secondary structure model of *A. aeolicus* tmRNA-TLD (left panel) and *E. coli* TLD (right panel) based on probing data and showing the probing results. Triangles are for V_1 cuts ; green arrows are S_1 cuts ; red nucleotides are cut by lead acetate.

Figure 3

Assignment of *A. aeolicus* tmRNA-TLD : Spectra recorded on a 1.2 mM sample of the *A. aeolicus* TLD, at 290K and pH 6.5. The T-stem and T and D- loop residues appear in green, the acceptor stem residues in red. A and B: extended contour plots of a 150 ms JR-NOESY experiment recorded at 600 MHz. The assignment walks is in solid green lines for T- stem and D and T- loop residues, and in solid red lines for the acceptor stem residues. The blue star

indicates the anticodon connectivity. Yellow circles emphasize the interloop cross-peaks. Crosspeaks between A42:T38 and G12:Ψ39. C: ^{15}N-^{1}H SG-TROSY experiment recorded in at 600 MHz and 290K. Same color coding as above.

Figure 4

TΨC – D loop interaction : *A. aeolicus* TLD assignment walks in the imino proton - aromatic proton region of the JR-NOESY spectra recorded at pH 6.5, 290K and 600 MHz. The color coding is as follows: purple for the adenine of the reversed-Hoogsteen A:T pair, green for the T and the Ψ of the TΨC loop, black for the T-stem. Crosspeaks between pairs A42:T38 and G12:Ψ39 are boxed

Figure 5

Assignment of *E. coli* tmRNA-TLD : Spectra recorded on a 1.5 mM sample of the *E. coli* TLD, at 290K and pH 6.5, using same color-coding as in Figure 3. Tentative assignment of T39NH and of the adjacent A:Ud pair (d for loop D) is drawn in light blue. A, B and C: extended contour plots of a 150 ms JR-NOESY experiment recorded at 600 MHz. The two arrows point to two aromatic resonances NOE-connected to T39CH3. D: ^{15}N - ^{1}H- JRSE-HMQC experiment recorded in natural abundance at 600 MHz and 290K.

Figure 6

Temperature titration at pH 6.5 of the imino proton region of (A) the *A. aeolicus* and (B) the *E. coli* TLDs. Spectra were recorded in the absence of salt and divalent cations.

Figure 7

Lower stack plot: 1D imino proton region of a 2.5 mM $MgCl_2$ *E. coli* TLD sample recorded at 293K and displayed *vs*. pH. Black bullets signal the apparition of a new structure at low pH. Black squares point some of the resonances corresponding to the structure observed at pH 6.5.

Upper strip plot: imino–methyl expanded region of a 130 ms JR-NOESY spectrum recorded at pH 4.55 and 283 K.

Figure 8

Models of the D- and TΨC- loop stacking compatible with the NMR data. The solid lines symbolises the sugar-phosphate backbone. T– stem and T-loop are in green, D- loop in red. (a) *A. aeolicus*. (b) *E. coli* at pH 6.5. The extra A:U base pair is colour-filled (c) Standard D-TΨC loop interaction in tRNAs.

Table 1. Alanylation properties of *in vivo* purified tmRNA-TLD, compared with those of *in vivo* purified full-length tmRNA.

RNAs	Charging level,* apparent (%)	K_M (μM)	V_{max} (nM s^{-1})	k_{cat} (min^{-1})	$G^{\#}$
tmRNA	30 ±15	9.5 ± 1.5	0.50 ± 0.05	0.25 ± 0.09	1
tmRNA-TLD	60 ±30	2.0 ± 0.2	1.0 ± 0.4	0.50 ± 0.07	9.5

* Charging levels of *E. coli* tmRNA-TLD (four independent experiments) *versus* full-length tmRNA (eleven independent experiments) were obtained for 15 and 30 min incubation with equivalent amounts of purified alanyl-tRNA synthetase from *E. coli*, respectively. All aminoacylations were performed at 37°C, as described.

$^{\#}$ Relative gain in catalytic efficiency (***G***) of tmRNA-TLD compared to tmRNA full-length, expressed as $\boldsymbol{G} = (k_{cat}/K_M)_{\text{tmRNA or tmRNA-TLD}} / (k_{cat}/K_M)_{\text{tmRNA}}$. Kinetic parameters represent an average of three experiments for tmRNA-TLD and of four experiments for full-length tmRNA, using independent enzyme purification's.

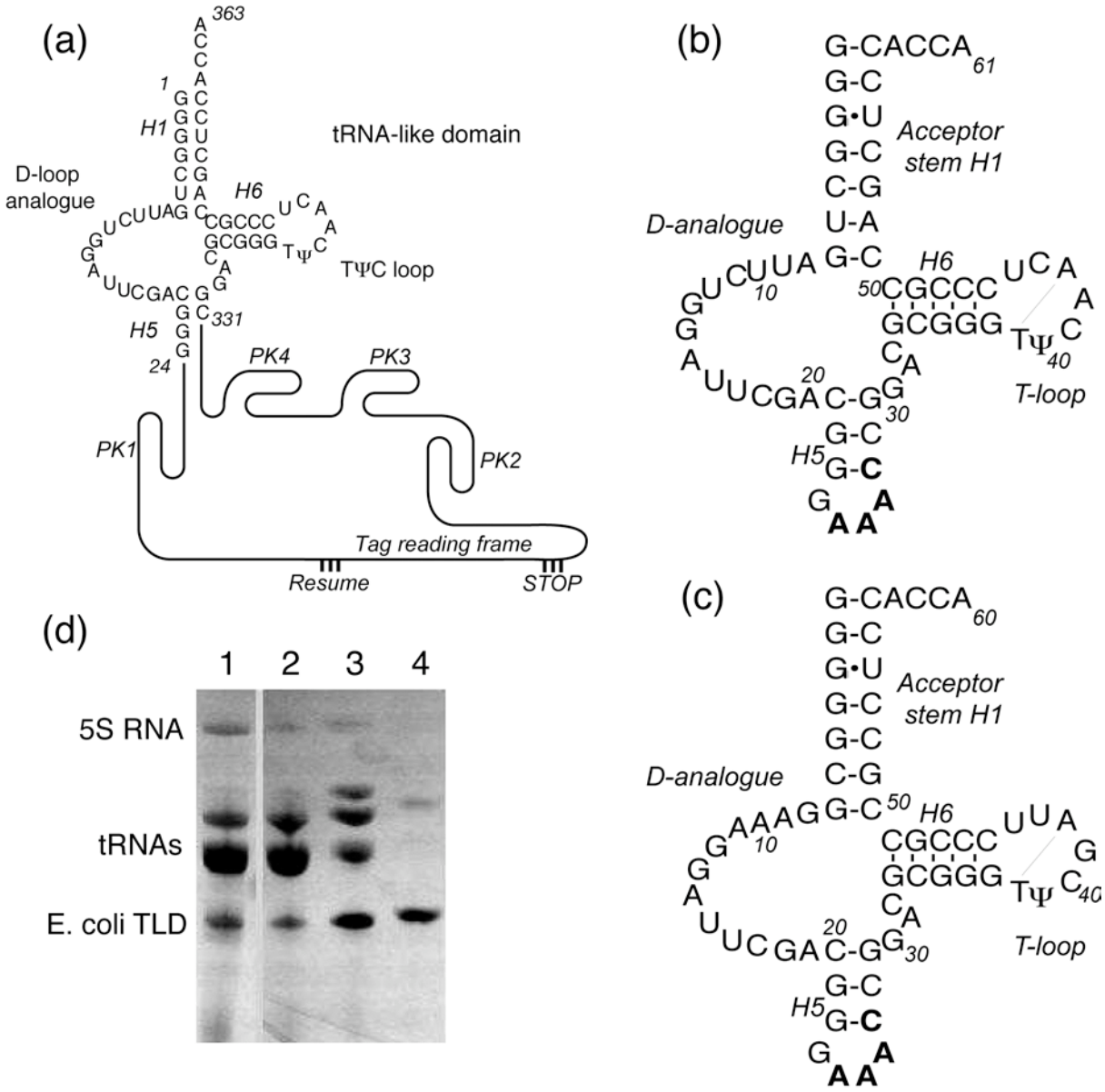
(a)
tRNA-like domain
D-loop analogue
TΨC loop
H1
H5
H6
PK1
PK2
PK3
PK4
Tag reading frame
Resume
STOP
(b)
Acceptor stem H1
D-analogue
T-loop
(c)
Acceptor stem H1
D-analogue
T-loop
(d)
1 2 3 4
5S RNA
tRNAs
E. coli TLD

Figure 1

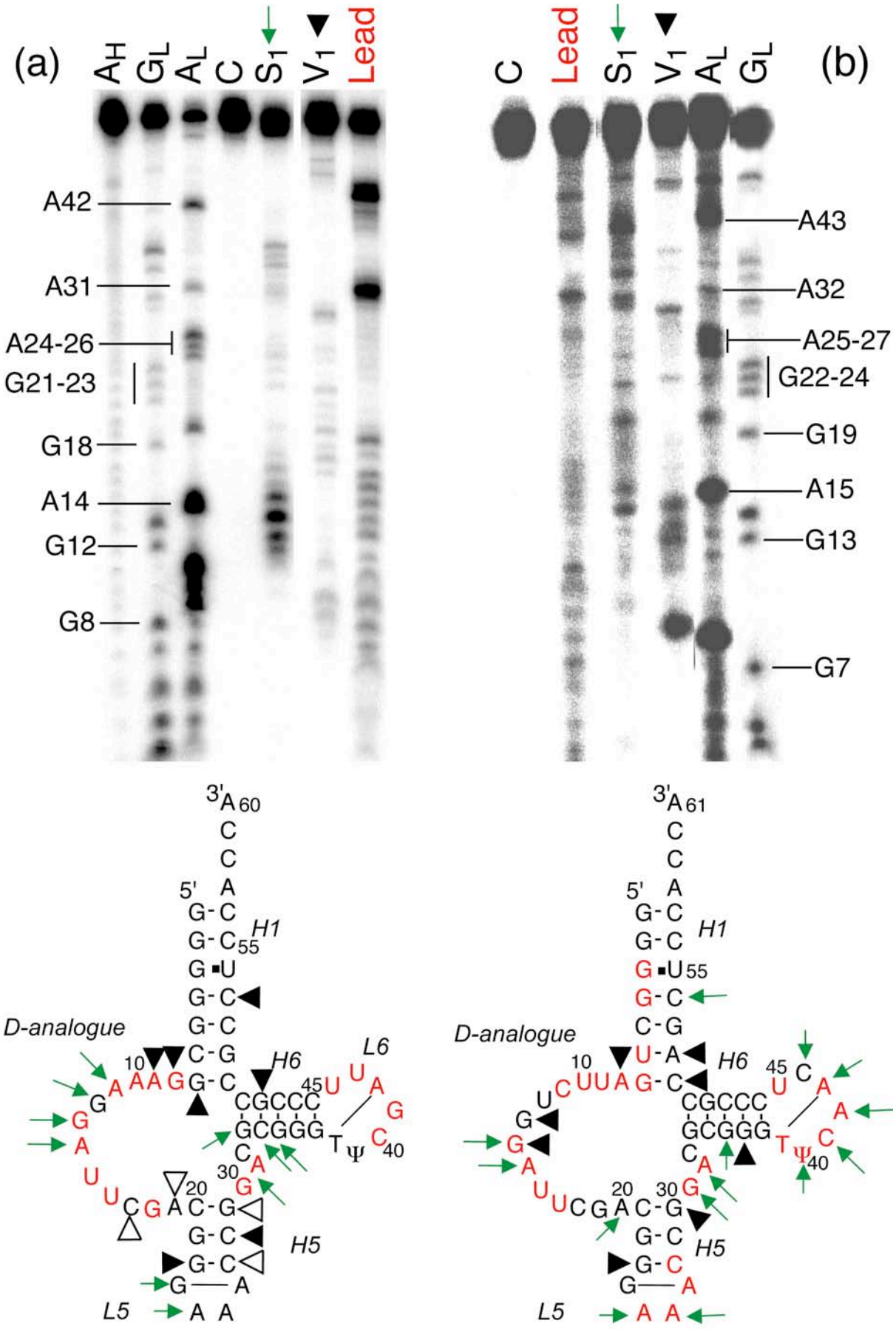
(a)
A_H
G_L
A_L
C
S1
V1
Lead
A42
A31
A24-26
G21-23
G18
A14
G12
G8
C
Lead
S1
V1
A_L
G_L
(b)
A43
A32
A25-27
G22-24
G19
A15
G13
G7
3'A 60
5'
H1
D-analogue
H6
L6
H5
L5
3'A 61
5'
H1
D-analogue
H6
H5
L5

Figure 2

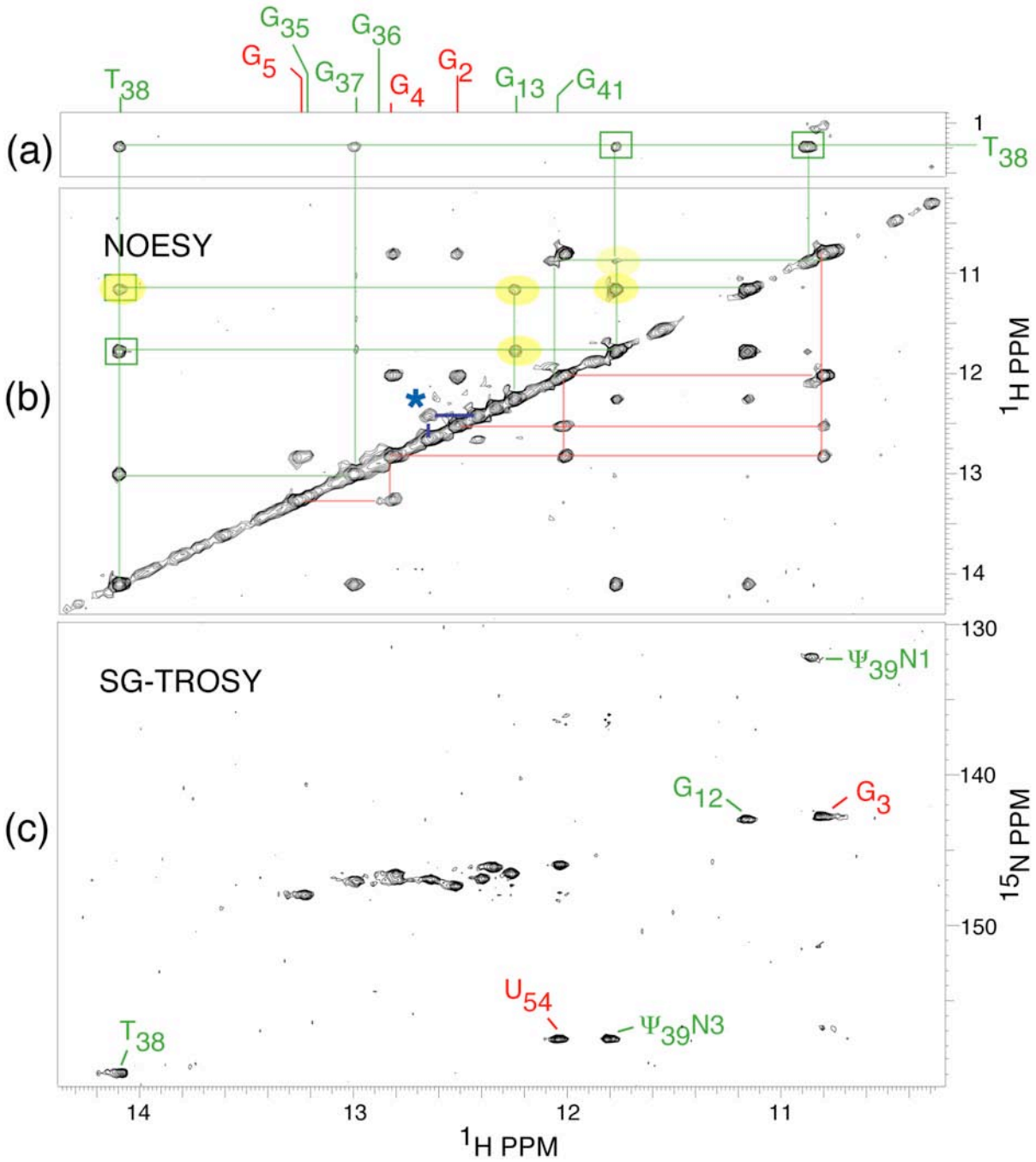
(a)
(b)
(c)
NOESY
SG-TROSY
T38
G35
G36
G5
G37
G4
G2
G13
G41
1
11
12
13
14
130
140
150
1H PPM
15N PPM
Ψ39N1
G12
G3
U54
Ψ39N3
T38
14
13
12
11
1H PPM

Figure 3

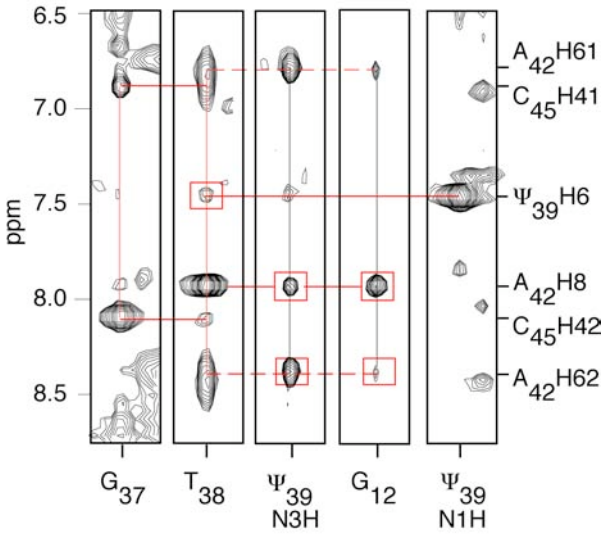
6.5
7.0
7.5
8.0
8.5
ppm
A42H61
C45H41
Ψ39H6
A42H8
C45H42
A42H62
G37
T38
Ψ39 N3H
G12
Ψ39 N1H

Figure 4

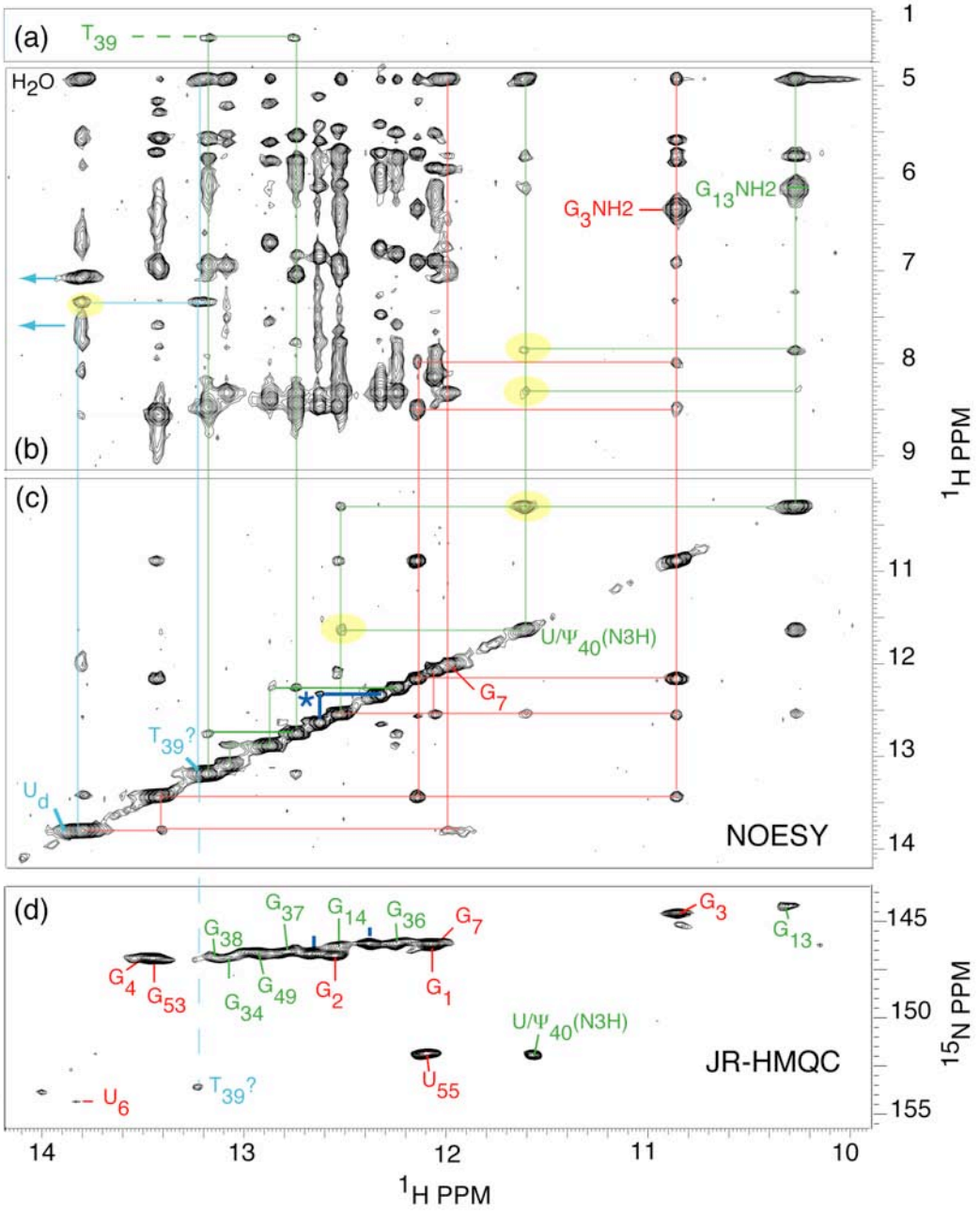
(a)
T39
H2O
G3NH2
G13NH2
(b)
(c)
U/Ψ40(N3H)
G7
T39?
Ud
NOESY
(d)
G37
G14
G36
G7
G38
G3
G13
G4
G53
G34
G49
G2
G1
U/Ψ40(N3H)
JR-HMQC
U55
T39?
U6
1H PPM
15N PPM

Figure 5

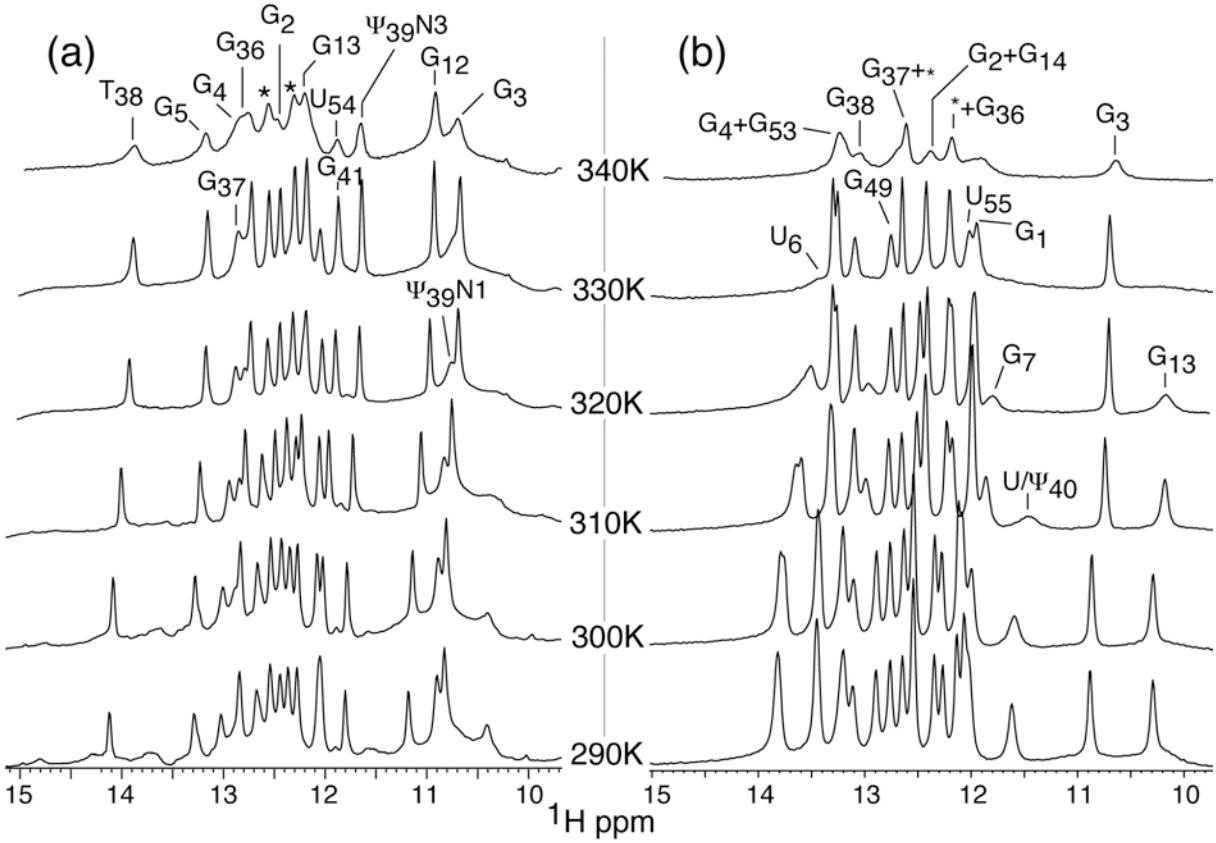
(a)
T38
G5
G4
G36
G2
*
*
G13
U54
Ψ39N3
G12
G3
G37
G41
Ψ39N1
(b)
G4+G53
G38
G37+*
G2+G14
*+G36
G3
G49
U55
G1
U6
G7
G13
U/Ψ40
340K
330K
320K
310K
300K
290K
15
14
13
12
11
10
1H ppm

Figure 6

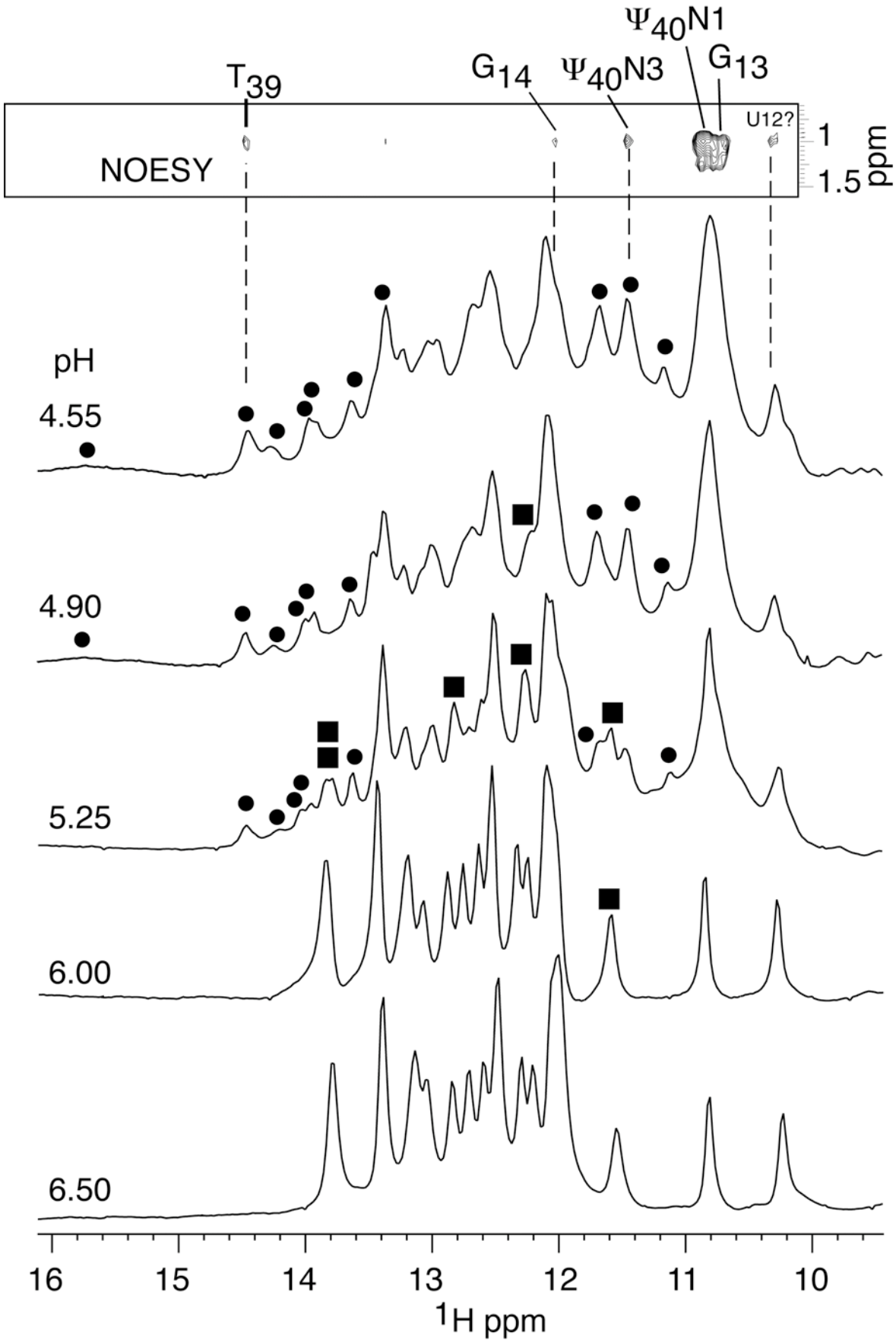
Ψ40N1
G13
G14
Ψ40N3
T39
U12?
NOESY
1
1.5
ppm
pH
4.55
4.90
5.25
6.00
6.50
16
15
14
13
12
11
10
1H ppm

Figure 7

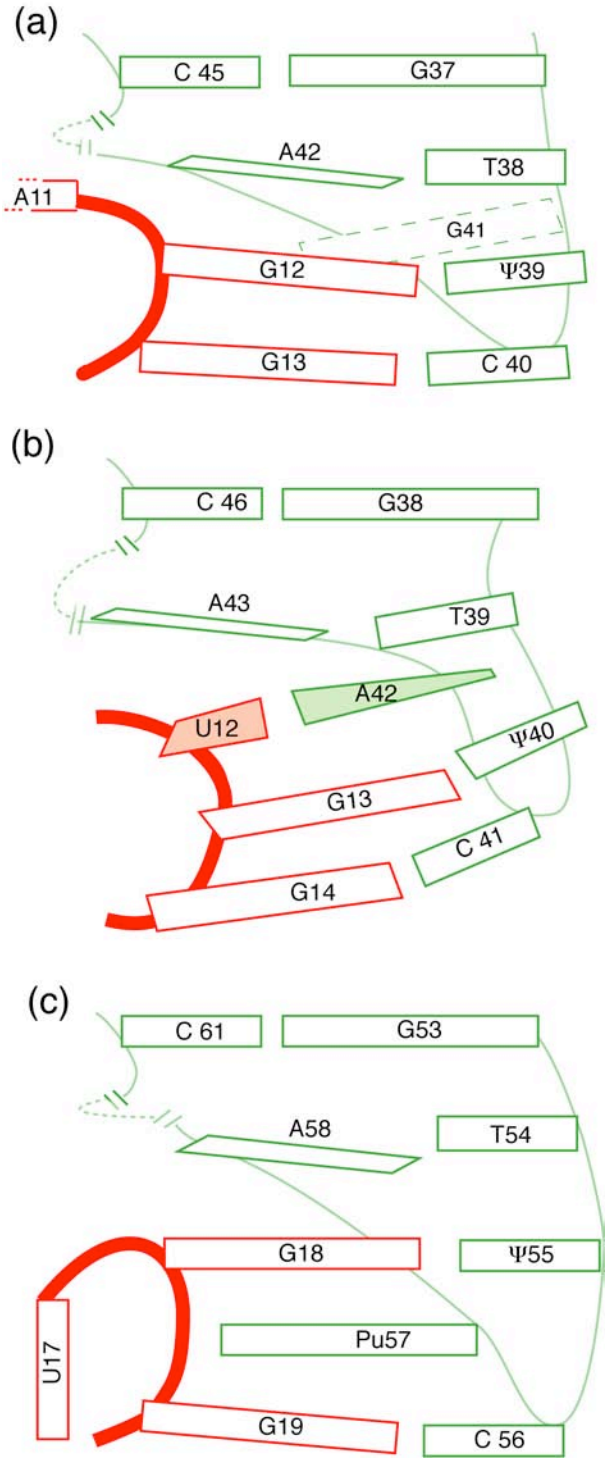
(a)
C 45
G37
A42
T38
A11
G41
G12
Ψ39
G13
C 40
(b)
C 46
G38
A43
T39
A42
U12
Ψ40
G13
C 41
G14
(c)
C 61
G53
A58
T54
G18
Ψ55
Pu57
U17
G19
C 56

Figure 8